Excerpts on Balanced Assessment

1.
Assessments *of* and *for* learning are both important. Since we in the U.S. already have many assessments *of* learning in place, if we are to balance the two, we must make a much stronger investment in assessment *for* learning. We can realize unprecedented gains in achievement if we turn the current day-to-day classroom assessment process into a more powerful tool for learning. We know that schools will be held accountable for raising test scores. Now we must provide teachers with the assessment tools needed to do the job.

It is tempting to equate the idea of assessment *for learning* with our more common term, "formative assessment." But they are not the same. Assessment *for learning* is about far more than testing more frequently or providing teachers with evidence so that they can revise instruction, although these steps are part of it. In addition, we now understand that assessment *for learning* must involve students in the process. When they assess *for learning*, teachers use the classroom assessment process and the continuous flow of information about student achievement that it provides in order to advance, not merely check on, student learning.

Assessment Crisis: The Absence of Assessment *FOR* Learning
Richard Stiggins, Phi Delta Kappan, 2002

2.
This report addresses assessments used in both classroom and large-scale contexts for three broad purposes: to assist learning, to measure individual achievement, and to evaluate programs. The purpose of an assessment determines priorities, and the context of use imposes constraints on the design. *Thus it is essential to recognize that one type of assessment does not fit all*.

Often a single assessment is used for multiple purposes; in general, however, the more purposes a single assessment aims to serve, the more each purpose will be compromised. For instance, many state tests are used for both individual and program assessment purposes. This is not necessarily a problem, as long as assessment designers and users recognize the compromises and trade-offs such use entails.

Although assessments used in various contexts and for differing purposes often look quite different, they share certain common principles. One such principle is that assessment is always a process of reasoning from evidence. By its very nature, moreover, assessment is imprecise to some degree. Assessment results are only estimates of what a person knows and can do.

Every assessment, regardless of its purpose, rests on three pillars: a model of how students represent knowledge and develop competence in the subject domain, tasks or situations that allow one to observe students' performance, and an interpretation method for drawing inferences from the performance evidence thus obtained. In the context of large-scale assessment, the interpretation method is usually a statistical model that characterizes expected data patterns, given varying levels of student competence. In less formal classroom assessment, the interpretation is often made by the teacher using an intuitive or qualitative rather than formal statistical model.

Knowing What Students Know: The Science and Design of Educational Assessment
Committee on the Foundations of Assessment, National Research Council, 2001

Excerpts on 21st Century Skills

1.

While the current assessment landscape is replete with assessments that measure knowledge of core content areas such as language arts, mathematics, science and social studies, there is a comparative lack of assessments and analyses focused on 21st century skills. Current tests fall short in several key ways:

- The tests are not designed to gauge how well students apply what they know to new situations or evaluate how students might use technologies to solve problems or communicate ideas.
- While teachers and schools are being asked to modify their practice based on standardized test data, the tests are not designed to help teachers make decisions about how to target their daily instruction.
- Current testing systems are rarely designed to measure a school or district's contribution to learning from a student's first day until his or her last day.

We must move from primarily measuring discrete knowledge to measuring students' ability to think critically, examine problems, gather information, and make informed, reasoned decisions while using technology. In addition to posing real world challenges, such assessments should accept a range of solutions to a task. For example, one possible assessment of 21st century skills would focus more on a student's operational skills, such as her expertise in using multiple sources appropriately and efficiently, rather than on whether or not a correct response was submitted.

21st Century Skills Assessment
Partnership for 21st Century Skills, 2007

2.

In an age when facts can be located in seconds on the Internet, schools should teach students such skills as how to critically evaluate what they find, connect it to other subjects, apply it to real-world problems, and collaborate well with partners halfway around the globe, advocates argue. Advances in psychology have shown that students are better engaged, and learn basic skills better, when they are taught in the context of more-complicated ones, such as solving a real-world problem.

Business and higher education leaders are pleading with schools to teach "21st-century skills," such as interdisciplinary thinking, that students need to flourish in an increasingly global, technology-rich society. But figuring out whether the skills have been taught well will require assessing them well, and that won't be easy or cheap.

Measuring such skills will require instruments more complex than multiple-choice tests, a prospect that sparks skepticism about whether the approach is feasible in a time of accountability, when some states are eliminating even short-answer items from tests to save time and money. [However,] new technology is facilitating the assessment of such skills.

Using more-complex tests raises significant issues, the Education Sector report says. One is a question of subjectivity when people, instead of machines, are required to score them. Another is the cost of training and monitoring scorers, and the time spent in scoring the tests.

Assessing '21st-Century Skills' Won't Be Easy, Paper Says
Catherine Gewertz, Education Week, 11/12/08

Excerpts on High School Assessment

1.
Results of a new study provide empirical evidence that, whether planning to enter college or workforce training programs after graduation, high school students need to be educated to a comparable level in reading and mathematics. The study results convey an important message to US high school educators and high school students: we should be educating all high school students according to a common academic expectation, one that prepares them for both postsecondary education and the workforce. Only then—whether they are among the two-thirds who enter college directly after graduation or those who enter workforce training programs—will they be ready. Recommended action steps that state policymakers can take:

- Require that all students take a rigorous core prep course program in high school.
- Hold schools and states accountable for preparing all students for college and workforce programs through rigorous core courses
- Provide funding for measures of college and workforce readiness skills to be used as statewide high school assessments.
- Begin measuring student progress with aligned assessments as early as the eighth grade.
- Communicate the common expectation of college and workplace training readiness to all stakeholders, including businesses, workforce and economic development associations, and educational institutions.

Ready for College and Ready for Work
ACT, 2006

2.
States are increasingly adopting end-of-course assessments for state accountability purposes or to satisfy NCLB assessment requirements in high school. In 2007, 18 states offered end-of-course assessments (Achieve, Policy Brief). The growing number of end-of-course assessments are designed to link high school assessments more closely with curriculum standards and courses students are required to take to graduate. End-of-course assessments also allow states to monitor rigor and consistency in courses taught statewide by offering a statewide common assessment, which is critical in states with high graduation standards.

Currently, eight states use end-of-course assessments in reading/language arts or mathematics for calculating AYP at the high school level. Most of these states require a single end-of-course assessment to be taken by all students (e.g., Algebra I). Other states require more than one end-of-course assessment for all students. For example, one state requires all students to take both an Algebra I and Geometry end-of-course assessment by the time the student graduates from high school. The results from both assessments are used to calculate AYP. Several additional states offer end-of-course assessments but do not include the results in AYP determinations.

End-of-Course Assessments
US Department of Education, National Technical Advisory Council, 9/08

Excerpts on Online Testing

1.
"I don't think any state has found it easy to go to computer-based testing," said Stuart R. Kahl, the president and chief executive officer of Measure Progress, a Dover, N.H.-based nonprofit organization that provides assessments to school districts. Although many districts have implemented computer-based testing at the district and school levels, using computers for statewide assessments is much more difficult, Mr. Kahl said.

One big difference between using computer-based tests at the local level and the state level is the limited testing window required for statewide assessments, Mr. Kahl pointed out. To keep students from sharing or discussing questions on the assessment, it's important for all students to complete the test within a very short time frame. But with such a narrow window, schools need more computers so students can take the test at the same time, and beefed-up network capacity to handle the spike in bandwidth required for the assessment applications to run smoothly. And even if schools have the hardware and the bandwidth to support computer-based statewide assessments, it doesn't mean they're ready to administer the tests, Mr. Kahl said. "There's so much more that has to happen."

Still, the advantages of computer-based assessments—quicker feedback, better-organized data, and a much less labor-intensive administration than with paper-and-pencil tests—are encouraging states to find solutions to those obstacles.

States Slow to Embrace Online Testing
Katie Ash, Education Week, 11/19/08

2.
In their plans, states need to distinguish between moving to an online administration of their testing program and adapting a full [Computer Adaptive Testing] CAT model. Both have important advantages and disadvantages, but these are not identical. The most significant differences between the two are cost, standards coverage, and NCLB compliance.

CAT models tend to be significantly more expensive than comparably sized computer-administered assessments for several reasons, including the sophistication of the adaptive engine required to support CAT and the much larger number of items needed to achieve the full CAT advantage.

Non-CAT computer-administered assessments are as likely to achieve high levels of alignment to content standards as their paper-based counterparts. CAT models must be designed specifically to do so or they will stop assessing a specific standard (or overassess it) when or until sufficient measurement accuracy is achieved. In fact, requiring CAT models to fully sample across the entire test blueprint may lessen CAT's primary advantage—efficiency of assessment time—because items will be included that no longer add to the reliability of the measurement.

Because of these breadth and depth standards coverage issues, proposed state CAT models have been rejected unless two conditions (not advantageous to CAT) are met: full breadth and depth of content coverage and inclusion of only grade-level content in the final score for each student.

Overview of Selected State Assessment Systems
Great Lakes West Comprehensive Center at Learning Point Associates, 1/09

Excerpts on Formative Assessment

1.

Formative assessment is a systematic process to continuously gather evidence about learning. The data are used to identify a student's current level of learning and to adapt lessons to help the student reach the desired learning goal. In formative assessment, students are active participants with their teachers, sharing learning goals and understanding how their learning is progressing, what next steps they need to take, and how to take them.

Formative assessment involves a variety of strategies for evidence gathering, which can be categorized into three broad types: on-the-fly assessment, planned-for interaction, and curriculum-embedded assessment.

- *On-the-fly assessment.* On-the-fly assessment occurs spontaneously during the course of a lesson. For example, a teacher listening to group discussions hears students expressing misconceptions about the science concept she has been teaching. She then changes the direction of her lesson to provide a quick "pop-up" lesson. The pop-up lesson enables the teacher to clear up the misconceptions before proceeding with her planned instructional sequence.
- *Planned-for interaction.* In planned-for interaction, teachers decide beforehand how they will elicit students' thinking during the course of instruction. For example, teachers plan the questions they will ask during the course of the lesson in order to enable students to explore ideas, and these questions can elicit valuable assessment information.
- *Curriculum-embedded assessments.* There are two kinds of curriculum-embedded assessments, those that teachers and curriculum developers embed in the ongoing curriculum to solicit feedback at key points in a learning sequence and those that are part of ongoing classroom activities. For example, student mathematical representations created during lessons can function as formative assessments, as can students' science notebooks that are also part of students' regular classroom activity.

Formative Assessment: What Do Teachers Need to Know and Do?
Margaret Heritage, Phi Delta Kappan, 10/07

2.

It's time to moderate our obsessions with standards, assessments, and accountability systems. We now need to shift our focus to providing teachers and administrators with appropriate and effective PD. Clearly, professional development is the best vehicle for changing instruction moving toward better formative assessment practices.

[W]hen it comes to diagnosing student learning, we need to start with teachers themselves. Quick fixes don't do the trick; teachers need help to hone the assessment skills they already have and guidance in appropriate ways to use external assessment tools. Well-designed professional development is the key.

Ongoing teacher and administrator training that allow educators to apply what they learn to improve their own practices is formative in and of itself. Year-long institutes, principal coaching models, and other approaches that sustain this support are the only ways to significantly change what's happening in many classrooms—and that's where achievement levels are raised.

Various Articles & Commentary from Test Vendor
Stuart Kahl, Measured Progress, Education Week Commentary

Excerpts on Benchmark Assessment

1.

Many have hoped that end-of-year [summative] tests would provide instructionally useful information for educators, they do not. This is not because there is something "wrong" with these summative accountability tests, rather that they were not designed to meet instructional purposes. Recognizing the inherent limitations of summative assessment, educators are looking for additional assessments to inform and track student learning during the year (generally termed "benchmark," "diagnostic" or "interim").

Interim assessment is the term we suggest for the assessments that fall between formative and summative assessment, including the medium-scale, medium-cycle assessments currently in wide use. Interim assessments (1) evaluate students' knowledge and skills relative to a specific set of academic goals, typically within a limited time frame, and (2) are designed to inform decisions at both the classroom and beyond the classroom level, such as the school or district level. Thus, they may be given at the classroom level to provide information for the teacher, but unlike true formative assessments, the results of interim assessments can be meaningfully aggregated and reported at a broader level. As such, the timing of the administration is likely to be controlled by the school or district rather than by the teacher.

These assessments may serve a variety of purposes, including predicting a student's ability to succeed on a large-scale summative assessment, evaluating a particular educational program or pedagogy, or diagnosing gaps in a student's learning.

The Role of Interim Assessments in a Comprehensive Assessment System: A Policy Brief
Marianne Perie, Scott Marion, Brian Gong, Judy Wurtzel; Achieve Inc., Aspen Institute, Center for Assessment 11/07

2.

Freeport Intermediate School (Texas) attributes its success to an unrelenting focus on results. Teachers work in collaborative teams for 90 minutes daily to clarify the essential outcomes of their grade levels and courses and to align those outcomes with state standards. They develop consistent instructional calendars and administer the same brief assessment to all students at the same grade level at the conclusion of each instructional unit, roughly once a week.

Each quarter, the teams administer a common cumulative exam. Each spring, the teams develop and administer practice tests for the state exam. Each year, the teams pore over the results of the state test, which are broken down to show every teacher how his or her students performed on every skill and on every test item. The teachers share their results from all of these assessments, and learn when a teammate has been particularly effective in teaching a certain skill. Team members consciously look for successful practice and attempt to replicate it; they also identify areas of the curriculum that need more attention.

Freeport Intermediate has been transformed from one of the lowest-performing schools in the state to a national model for academic achievement. Principal Clara Sale-Davis believes that the crucial first step in that transformation came when the staff began to honestly confront data on student achievement and to work together to improve results.

What Is a "Professional Learning Community"?
Richard DuFour, Schools as Learning Communities , Vol 61, May 2004